MATH 343 Lec 22

Let w be the binary vector which is manipulated or controlled by the experimenter. The data looks like:

$$\mathbb{D} = \left\langle \left[\vec{w} \,|\, \vec{x}_{\cdot 1} \,|\, \vec{x}_{\cdot 2} \,|\, \ldots \,|\, \vec{x}_{\cdot p} \right], \vec{y} \right\rangle$$ where $\vec{x}_{\cdot j}$'s are called "baseline measurements" measured before the response $\vec{y}$. We will assume they're known at the time we manipulate $\vec{w}$ (non-sequential ~~experiment~~ setting).

The vector $\vec{w}$ is called the "assignment" or "allocation" since you are assigning / forcing / manipulating subject i to receive w_i.

How do we generate $\vec{w}$'s? There are 2^n possibilities! The means of generating $\vec{w}$'s is called the "design". What design is "best"?

Difficult problem! My whole research program of the past 10yr focuses on it!

To measure "best" we need to focus on specific settings and make assumptions.

Setting: $\mathcal{Y} = \mathbb{R}$ (regression), n subjects, $w_i \in \{0,1\}\ \forall i$.
for now, no other subject measurements exist. $X = [\vec{1}_n \mid \vec{w}]$.
We assume the "population model":

let $\vec{Y} = \beta_0 \vec{1}_n + \beta_T \vec{w} + \vec{\mathcal{E}}$ where $\mathcal{E}_1, \ldots, \mathcal{E}_n \overset{iid}{\sim}$ with mean 0, variance σ^2
(homoskedastic errors).

$\Rightarrow E_{\vec{\mathcal{E}}}[\vec{Y}] = \beta_0 \vec{1}_n + \beta_T \vec{w} \Rightarrow E_{\mathcal{E}_i}[Y_i] = \beta_0$ if $w_i = 0$ and $E_{\mathcal{E}_i}[Y_i] = \beta_0 + \beta_T$ if $w_i = 1$

$\Rightarrow \beta_T$ is the causal population additive treatment effect (PATE), our target parameter of interest.

let $n_T := \sum_{i=1}^n w_i$, the # of subjects assigned to treatment, $p_T := \frac{n_T}{n}$, the prop.
let $n_C := n - n_T$

let $\vec{B} := \begin{bmatrix} B_0 \\ B_T \end{bmatrix} = (X^\top X)^{-1} X^\top \vec{Y}$, the OLS estimator

$E_{\vec{\mathcal{E}}}[B_T] = \beta_T \Rightarrow$ unbiased $\Rightarrow MSE_{\vec{\mathcal{E}}}[B_T] = Var_{\vec{\mathcal{E}}}[B_T] = \sigma^2 (X^\top X)^{-1}_{2,2}$

Let's design $\vec{w}$ to minimize variance of B_T:

$$\vec{w}_* := \underset{\vec{w} \in \{0,1\}^n}{\text{argmin}} \left\{ \sigma^2 (X^\top X)^{-1}_{2,2} \right\} = \text{argmin} \left\{ \frac{\sigma^2}{n} \frac{1}{1-p_T} \begin{bmatrix} 1 & -1 \\ -1 & p_T^{-1} \end{bmatrix}_{2,2} \right\} = \text{argmin} \left\{ \frac{1}{1-p_T} p_T^{-1} \right\}$$

$$= \text{argmax} \{ p_T (1-p_T) \}$$

$$= \{ \vec{w} : p_T = \tfrac{1}{2} \}$$

$$X^\top X = [\vec{1}_n \mid \vec{w}]^\top [\vec{1}_n \mid \vec{w}] = \begin{bmatrix} \leftarrow \vec{1}_n \rightarrow \\ \leftarrow \vec{w}_n \rightarrow \end{bmatrix} \begin{bmatrix} \uparrow & \uparrow \\ \vec{1}_n & \vec{w} \\ \downarrow & \downarrow \end{bmatrix} = \begin{bmatrix} n & n_T \\ n_T & n_T \end{bmatrix} = n \begin{bmatrix} 1 & p_T \\ p_T & p_T \end{bmatrix}$$

$$(X^\top X)^{-1} = \frac{1}{n} \frac{1}{p_T - p_T^2} \begin{bmatrix} p_T & -p_T \\ -p_T & 1 \end{bmatrix} = \frac{1}{n} \frac{1}{1-p_T} \begin{bmatrix} 1 & -1 \\ -1 & p_T^{-1} \end{bmatrix}$$

$$Var[B_T] = \frac{\sigma^2}{n} \frac{1}{1-\frac{1}{2}} (2) = \frac{4\sigma^2}{n}$$

As long as $p_T = \frac{1}{2} \Leftrightarrow n_T = n_C$ called "equal allocation", we have the optimal design.

What is the estimator under equal allocation?

$$\vec{B} = (X^\top X)^{-1} X^\top \vec{Y} = \frac{1}{n} \frac{1}{1-p_T} \begin{bmatrix} 1 & -1 \\ -1 & p_T^{-1} \end{bmatrix} \begin{bmatrix} \vec{1}^\top \vec{Y} \\ \vec{w}^\top \vec{Y} \end{bmatrix} = \frac{1}{(\frac{n}{2})} \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} \vec{1}^\top \vec{Y} \\ \vec{w}^\top \vec{Y} \end{bmatrix} = \frac{1}{(\frac{n}{2})} \begin{bmatrix} \vec{1}^\top \vec{Y} - \vec{w}^\top \vec{Y} \\ 2\vec{w}^\top \vec{Y} - \vec{1}^\top \vec{Y} \end{bmatrix}$$

let $I_T := \{ i : w_i = 1 \}$,
let $I_C := \{ i : w_i = 0 \}$

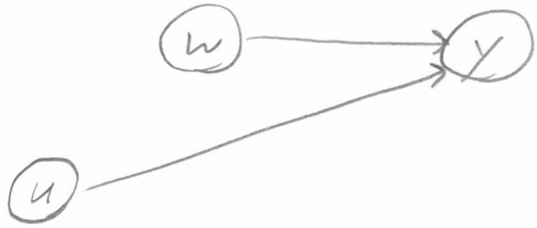
$$\Rightarrow B_T = \frac{2\vec{w}^\top \vec{Y} - \vec{1}^\top \vec{Y}}{\frac{n}{2}} = \frac{2 \sum_{\{i : w_i = 1\}} Y_i - \sum Y_i}{\frac{n}{2}} = \frac{2 \sum_{i \in I_T} Y_i - \left(\sum_{i \in I_T} Y_i + \sum_{i \in I_C} Y_i \right)}{\frac{n}{2}} = \frac{\sum_{i \in I_T} Y_i - \sum_{i \in I_C} Y_i}{\frac{n}{2}} = \bar{Y}_T - \bar{Y}_C$$

"default" "simple" estimator

Does it matter which subjects get assigned $w_i=1$ or $w_i=0$ specifically? NO!

So can we assign $\vec{w} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}\begin{matrix} \uparrow \\ n_T \\ \downarrow \\ \uparrow \\ n_C \\ \downarrow \end{matrix}$ Sure!!

But now let's consider a slightly different situation: one unobserved measurement, u, and a ~~DAG~~ as follows with structural model as follows



where $\vec{Y} = \beta_0 \vec{1}_n + \beta_u \vec{u} + \beta_T \vec{w} + \vec{\mathcal{E}}$, $\mathcal{E}_1, \ldots \mathcal{E}_n \stackrel{iid}{\sim}$ mean zero, variance σ^2

and we use the simple estimator and equal allocation

$$b_T = \bar{Y}_T - \bar{Y}_C = \frac{1}{\frac{n}{2}} \sum_{i \in I_T} \beta_0 + \beta_u u_i + \beta_T + \mathcal{E}_i - \frac{1}{\frac{n}{2}} \sum_{i \in I_C} \beta_0 + \beta_u u_i + \mathcal{E}_i$$

$$= \frac{1}{\frac{n}{2}} \left(\left(\frac{n}{2}\beta_0 + \beta_u \sum_{i \in I_T} u_i + \frac{n}{2}\beta_T + \sum_{i \in I_T} \mathcal{E}_i \right) - \left(\frac{n}{2}\beta_0 + \beta_u \sum_{i \in I_C} u_i + \sum_{i \in I_C} \mathcal{E}_i \right) \right)$$

$$= \beta_T + \beta_u(\bar{u}_T - \bar{u}_C) + (\bar{\mathcal{E}}_T - \bar{\mathcal{E}}_C)$$

$$\Rightarrow E_{\vec{\mathcal{E}}}[b_T] = \beta_T + \underbrace{\beta_u(\bar{u}_T - \bar{u}_C)}_{\text{Bias!}}$$

Right... but you don't know what the values of u_i are... so who cares? If they're random and unrelated to anything known, it doesn't matter! But...

What if you suspect u is a function of subject #... like increases monotonically

$\Rightarrow \bar{u}_T - \bar{u}_C$ is large if $\vec{w} = \begin{bmatrix} \vec{1}_{n_T} \\ \vec{0}_{n_C} \end{bmatrix}$. Or...

What if someone knew that $\vec{w} = \begin{bmatrix} \vec{1}_{n_T} \\ \vec{0}_{n_C} \end{bmatrix}$ so they order the subjects so the large u_i's come first $\Rightarrow$ biases estimate of β_T $\uparrow$.